Identifying Patients with Major Diabetes-related Complications

(N2C2 Shared Task 2018, Track 1)

Madhumita Sushil
Major diabetes-related complications

- Amputation
- Kidney Damage
- Skin conditions
- Retinopathy
- Neuropathy
- Nephropathy
Dataset

Set of ~200 Patients

Has any complication (criteria “met”) or not (criteria “not met”)
Approaches
Major complications: UMLS extension

- Amputation
- Kidney Damage
- Skin conditions
- Retinopathy
- Neuropathy
- Nephropathy

- Renal Damage
- Skin condition
- Skin disorder
- Dermatologic disorder
Major complications: CUI resolution from UMLS

- Diabetes
  - Amputation (C0002688)
  - Kidney Damage (C1408258)
  - Skin conditions
    - Retinopathy (C0011884)
    - Neuropathy (C0011882)
    - Nephropathy (C0011881)

- Skin condition (C1719933)
- Dermatologic disorder (C0037274)
Embeddings

Word embeddings:

- Word2vec embeddings trained on Pubmed
- “Complication embeddings” - average word vectors

CUI embeddings:

- Pretrained CUI embeddings (Beam et al., 2018)
## Concept-embeddings

<table>
<thead>
<tr>
<th>Concept recognition</th>
<th>Concept filtering</th>
<th>Inducing embeddings</th>
</tr>
</thead>
<tbody>
<tr>
<td>1) Mention</td>
<td>Limit to “problems” that are “present”</td>
<td>“Problem” embeddings</td>
</tr>
<tr>
<td>2) Semantic type</td>
<td></td>
<td>1) Average word vector embeddings</td>
</tr>
<tr>
<td>3) Assertion</td>
<td></td>
<td>2) CUI embeddings</td>
</tr>
<tr>
<td>4) CUI</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Matching criteria

1) Regex match to “diabetes”, “diabetic”, and “DM” to check if patient has diabetes, “not met” otherwise

2) If diabetic, and for any “present” problem complication pair

\[ \cos(\text{prob}_\text{emb}, \text{comp}_\text{emb}) > 0.65 : \text{criteria “met”} \]

Criteria “not met” otherwise
Combining CUI and word embedding similarity

Using both approaches and combining them:

1) Max of both similarity scores
2) Average to both similarity scores
# Results (train)

<table>
<thead>
<tr>
<th></th>
<th>met</th>
<th>not met</th>
<th>overall</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Prec.</td>
<td>Rec.</td>
<td>Speci.</td>
</tr>
<tr>
<td>Majority</td>
<td>0.5594</td>
<td>1.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td>RF - BoW</td>
<td>0.7767</td>
<td>0.7080</td>
<td>0.7416</td>
</tr>
<tr>
<td>NB - BoW</td>
<td>0.6827</td>
<td>0.6283</td>
<td>0.6292</td>
</tr>
<tr>
<td>RF - BoEmbCluster</td>
<td>0.6194</td>
<td>0.7345</td>
<td>0.4270</td>
</tr>
<tr>
<td>Word emb sim</td>
<td>0.8404</td>
<td>0.6991</td>
<td>0.8315</td>
</tr>
</tbody>
</table>
More Results (train)

<table>
<thead>
<tr>
<th></th>
<th>met</th>
<th>not met</th>
<th>overall</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Prec.</td>
<td>Rec.</td>
<td>Speci.</td>
</tr>
<tr>
<td>Word emb sim</td>
<td>0.8404</td>
<td>0.6991</td>
<td>0.8315</td>
</tr>
<tr>
<td>CUI similarity</td>
<td>0.8356</td>
<td>0.5398</td>
<td>0.8652</td>
</tr>
<tr>
<td>Max similarity</td>
<td>0.7807</td>
<td>0.7876</td>
<td>0.7191</td>
</tr>
<tr>
<td>Avg similarity</td>
<td>0.8020</td>
<td>0.7168</td>
<td>0.7753</td>
</tr>
</tbody>
</table>
Thank You!